Doing good online: An investigation into the characteristics and motivations of digital volunteers

Joe Cox[1a], Eun Young Oh[a], Brooke Simmons[b], Gary Graham[c], Anita Greenhill[d], Chris Lintott[b], Karen Masters[a], and Jamie Woodcock[c]

[a] *University of Portsmouth, Winston Churchill Avenue, Portsmouth, PO1 2UP, UK*

[b] *University of Oxford, Wellington Square, Oxford, OX1 2JD, UK*

[c] *Leeds University, Leeds, LS2 9JT, UK*

[d] *University of Manchester, Oxford Rd, Manchester, M13 9PL, UK*

ABSTRACT

The rapid growth in Internet technology is making it possible to volunteer in online settings, with participants able to contribute directly to research-based activities supporting non-profit groups and charitable organisations. This study undertakes an investigation into the profile and motivations of contributors to these online volunteering projects. We specifically investigate volunteer activity and retention for the online crowdsourcing platform known as the Zooniverse, which is home to around thirty online volunteering projects. Through a survey undertaken with a representative sample of contributors and reconciling against records of actual voluntary activity, we are able to measure motivations against the Volunteer Functions Inventory (VFI) and explore relationships with observed levels of activity and retention. Our results show that a unique combination of 'other' and 'self' oriented motivation, specifically Protective & Enhancement, Values and particularly Understanding, associate significantly and positively with observed variations in volunteering activity and retention in an online setting.

Keywords: Online; Volunteering; Motivations; Volunteer Functions Inventory

JEL Codes: A13; D12; L30; L86

[1] Corresponding author. Tel.: +44 23 9284 4723. *Email address:* joe.cox@port.ac.uk.

1: Introduction

While much has been written concerning the growth of the digital economy and its impact upon commercial and for-profit activities, relatively little attention has been paid to the effects of digitisation upon the voluntary sector. The sharp reductions in search and transactions costs associated with online interactions have profoundly affected the ways in which ordinary people are able to actively contribute towards socially valuable causes. Online volunteering projects are truly many and varied, but involve aggregation of inputs from very large numbers of contributors working together towards a common goal. Possibly the best-known among such projects is Wikipedia, an online encyclopaedia co-created and maintained exclusively by volunteer contributors with the aim of 'allowing free access to the sum of all human knowledge for everyone on the planet' (Wikipedia, 2015).

This new form of online volunteering may have significant implications for the academic study of voluntary activity, including our understanding of motivation. In the literature on human-computing interaction, Amichai-Hamburger (2008) has previously undertaken research into the United Nations 'Online Volunteering.org' initiative where volunteers contribute their skills to online projects to help with development issues, finding that participants tend to be motivated by the possibility of self-actualisation. Yang & Lai (2010) arrive at a similar conclusion in their study of intrinsic and extrinsic motivations among contributors to Wikipedia, finding internal self-concept based motivations to be the most prevalent. Dhebar & Stokes (2008) also find that regular communication between organisers and participants to be a key motivational factor behind participation in online volunteering assignments. However, outside of these few studies, an overwhelming majority of the existing theory and evidence on volunteer motivation is based on conventional 'real world'

activities, leading to a need to develop a more detailed understanding of the effects of digitisation on what is already known in relation to volunteer engagement.

This study makes a unique contribution to this body of literature by presenting and analysing results from a large-scale survey undertaken with a representative sample of registered users of the 'Zooniverse', a web-based portal that is home to around thirty online volunteering projects. Zooniverse projects allow citizens to participate in collaborative research activities and are managed by teams based in museums, universities and other non-profit or charitable organisations, such as Cancer Research UK, The Tate Gallery, The Imperial War Museum and the Gorongosa National Park. Our survey dataset contains information on socio-demographic, attitudinal and behavioural information consistent with the established Volunteer Functions Inventory (VFI). We reconcile this against an extensive database of user interactions recorded directly by the Zooniverse in order to examine to examine the extent to which VFI motivations can explain variations in activity and retention levels among individual volunteers. Much of our analysis is therefore based upon observed rather than stated behaviours, which contrasts with a vast majority of prior studies on volunteering that are limited by their reliance on self-reported activity levels.

This study specifies three specific research questions that we address through the analysis of these data. Our first research question involves an investigation into the profile of online volunteers to establish the extent to which they are representative both of the population as a whole and of volunteers more generally. Our second research question concerns the analysis of volunteer responses to items from the VFI and the extent to which motivations can be reliably and consistently measured in this way. To address this, we employ a Principal Component Analysis (PCA) alongside the raw response data grouped by the resultant categories to illustrate the commonly held motivations expressed by our sample of online

volunteers. Our third and final research question involves formally modelling volunteer engagement with Zooniverse projects through the estimation of a series of multiple regressions using a range of measures of volunteer activity and retention as dependent variables. Our independent variables include the set of factor scores relating to items from the VFI alongside other socio-demographic, attitudinal and behavioural controls captured by our survey data. Our results demonstrate a particular subset of motivations and characteristics which are able to explain variations in volunteer engagement with online volunteering projects.

The remainder of the paper is organised as follows. Section 2 presents a more detailed explanation of online volunteering and the Zooniverse platform in general, while Section 3 outlines the conceptual framework used in this study. A discussion and analysis of our unique data set and method is presented in Section 4, while Section 5 consists of separate sub-sections investigating our three core research questions. A summary and set of concluding remarks are finally presented in Section 6.

2: Online Volunteering and the Zooniverse

One of the best known voluntary crowdsourcing platforms in the field of non-commercial research is the Zooniverse, a collection of around thirty active online research projects powered by volunteer contributors (Fortson *et al.*, 2012). Zooniverse projects represent a unique response to challenges posed by increasingly large and visually complex data sets which cannot analysed using computer algorithms alone, but where humans are able to interpret much of the relevant information contained within the data. Zooniverse projects

therefore ask for input from volunteers to assess and classify large datasets with the objective of helping teams of professional researchers in non-profit organisations and charities to address a range of specific research questions. The first and one of the best-known Zooniverse sites, Galaxy Zoo, confronts users with a series of images of deep-space galaxies (Lintott *et al*., 2008) and asks them to classify these galaxies according to a set of pre-defined criteria relating to their shape and internal structure. The resultant analysis of data gathered from volunteers is helping astrophysicists develop a better understanding of the evolution of galaxies. Other examples of Zooniverse projects include Cell Slider, which asks volunteers to analyse the properties of cancer cells to help Cancer Research UK develop possible treatments; AnnoTate, where volunteers transcribe artists' notes held in the Tate collection and Wildcam Gorongosa, where volunteers classify animals appearing in images from camera traps stationed around the Gorongosa National Park[2]. While any individual contributor has the potential to misclassify the information they are asked to interpret, these projects are based on the collection of such information from a large number of independent assessments from different volunteers. This approach taps into the well-documented 'wisdom of crowds' phenomenon first noted by Galton (1907a; 1907b) in the context of a contest to guess the weight of an ox and has been more recently popularised by Surowiecki (2004).

The Zooniverse has been hugely successful since its launch in 2010 and now has around 1.3 million registered volunteers. On average across each Zooniverse project, volunteer workers contribute the amount of information that it would take a professional researcher 34 full-time working years to complete alone (Cox *et al*., 2015). The work of contributors to these projects is also making a significant contribution to society's knowledge and understanding

[2] A full list of current Zooniverse projects can be found at www.zooniverse.org.

on issues of key importance, while volunteers have even uncovered entirely new phenomenon through their participation. Most famously, Dutch school teacher Hanny Van Arkel discovered an entirely new astronomical phenomenon (Hanny's Voorwerp) that was previously unknown to science while volunteering for the Galaxy Zoo project (Lintott *et al*., 2009).

A very limited number of studies have looked at the motivations of contributors to Zooniverse projects. Raddick *et al.* (2010) investigated motivations among participants of the Galaxy Zoo project, first holding interviews with a smaller number of individual contributors to ask about their motivations and subsequently grouping these into discrete categories and surveying a larger sample of users in a follow-on study (Raddick *et al.*, 2013). They found that 'being excited by the opportunity to make an original contribution to science' was most commonly stated as the most important motivation. However, both of the above studies suffer from the high likelihood of selection bias in the composition of the sample and the absence of any investigation into the relationship between identified motivations and patterns of volunteer activity and retention. In addition to overcoming each of these limitations, another benefit of our study is that we do not simply investigate volunteering activity recorded for a single project. Instead, we gather data from volunteer contributors for a number of projects in the areas of astrophysics and ecology; specifically, Galaxy Zoo, Planet Hunters, Seafloor Explorer, Snapshot Serengeti and Penguin Watch. Figure 1 (below) contains screenshots of the online interface for each of these projects. In each case, a volunteer is either asked to answer a series of questions about the properties of an image they see, or are asked to point and click to areas of an image relating to content of particular research interest. Sophisticated algorithms are subsequently applied to convert the large quantity of volunteer data supplied for each individual image into a consensus solution which

can be used for research. The high quality of the research data generated by Zooniverse projects is highlighted by the 91 publications[3] in peer-reviewed academic journals that have only been possible as a result of input from online volunteers, many of whom are formally thanked in the author acknowledgements or credited as formal co-authors.

[Figure 1 about here]

3: Conceptual Framework

This study employs the functional approach to human behaviour in order to understand motivations to volunteer for online projects of the kind typified by the Zooniverse. This approach is largely based on theories of Smith *et al.* (1956) and Katz (1960) which assert that volunteers are motivated by a desire to satisfy various combinations of social and psychological goals, such as acquiring understanding, expressing important values, protecting the ego, forming social bonds and responding to rewards and punishments. The most well-known and complete metric used to measure and interpret volunteer motivation was pioneered by Clary *et al.* (1996) and is known as the Volunteer Functions Inventory (VFI); a formal instrument for measuring volunteer motivations consisting of six distinct items. These motivations are; Protective (a means to shield or escape from problems); Enhancement (a means to feel better about oneself); Social (a means to interact with people and expand social networks); Values (as a means to express personal values and contribute to causes identified

[3] A full list of all peer-reviewed publications resulting from Zooniverse projects can be found at https://www.zooniverse.org/publications

as being important); Understanding (a means to gain new perspectives and to learn) and Career (a means to build skills and connections to enhance one's career). These items have been shown to be robust and consistent when applied across different cohorts of volunteers, as well as across time and different forms of volunteering (Clary *et al*., 1998).

The seminal paper by Clary *et al.* (1996) utilised the VFI to identify motivations among respondents to a US survey on volunteering and giving, concluding that the Values, Career, Social and Understanding motivations tended to dominate among their sample, while the Protective and Enhancement motivations were not found to be particularly prevalent. Building on these findings, the VFI has subsequently seen widespread use in analysing the motivations of volunteers for a number of activities and organisations around the world. Although results tend to differ somewhat depending on the particular context, a majority of studies highlight the importance of the more 'other-oriented' motivations of Values, Understanding and Social, with lower importance attached to the more 'self-oriented' motivations of Protective, Enhancement and Career (Planalp & Trost, 2009; Agostinho & Paco, 2012).

Other studies have explored how the volunteer motivations conceptualised by the VFI could be used to predict volunteer behaviours, including recruitment and retention based on matching of motivation to particular volunteering contexts and environments (Clary & Snyder, 1999; Stukas *et al.*, 2009). Again, in common with the above, studies relating volunteer motivations to behaviours tend to find a positive association between frequency of volunteering and the Values and Understanding motivations (Allison *et al.*, 2002; Gage & Thapa, 2012; Stukas *et al.*, 2014). Other studies such as Garner & Garner (2011) and Misje *et al.* (2005) show that the longest serving volunteers are significantly more likely to express other-regarding motives (especially Values) compared with shorter-term volunteers who are

more likely to be motivated by self-regarding motivations. A number of more recent studies have begun to explore inter-relationships between different elements of the VFI. For example, Peachy *et al.* (2014) use the VFI to inform a series of interviews conducted with a sample of 'sport-for-development' volunteers, finding strong evidence of a strong positive relationship between the Understanding and Career motivations.

The only prior study of which we are aware that has applied the VFI in in the context of online volunteering did so in relation to Wikipedia contributors (Nov, 2007), finding that the more altruistic motivations of Values and Understanding were more prevalent and tended to do a better job of predicting variations in self-reported activity levels. However, this study is affected by a number of key limitations, given that the findings are based on analysis of data from a self-selecting group of survey respondents which is not necessarily representative of the population being studied, while also relying on testing relationships between motivations and self-reported activity levels. Our study overcomes these limitations as the result of gathering data from a large and representative sample of online volunteers, while also reconciling the survey data with actual recorded patterns of volunteer engagement recorded by the Zooniverse platforms. This means that our data are much less likely to be biased due to disproportionately high participation amongst the most actively engaged users and is also much more likely to establish accurate links between motivation and behaviour given that measures of participation are based on actual observed patterns rather than relying on self-reported data.

4. Data

A combination of two data sources are used as part of this study. A majority of our data come from a survey of Zooniverse users undertaken during April and May 2015, spanning five different individual projects. The survey was entirely web-based, with each individual respondent being e-mailed a unique URL that would enable us to link their responses to their Zooniverse user accounts. This allowed us to also collect information directly from the Zooniverse database so that each set of survey responses could be matched up against their historical patterns of activity, including the amount of data analysis (number of 'classifications') supplied, recorded for both the 'home' project (to which the Volunteer contributes most often) and aggregated across the entire portfolio of Zooniverse projects. We also capture amount of time actually spent classifying and the number of individual projects towards which the volunteer has contributed. We measure the retention of volunteers through the number of unique days/log-in sessions recorded and the length of time for which each respondent has 'actively' contributed towards projects (the time difference between first and last recorded classifications). This range of measures reflects a number of dimensions of volunteer activity and retention, while also checking for the robustness of our findings. After excluding a very small number of obvious outliers, our final dataset comprises a total of 1,915 respondents drawn from five different Zooniverse projects.

As with many voluntary web-based surveys, obtaining a representative sample of users was a primary concern given that we were otherwise likely to encounter a disproportionately active sample of volunteers. To overcome this, we launched a smaller pilot survey prior to the full release where we invited a randomly determined subset of users to participate. By comparing the number of responses with the number of invitees among users demonstrating different levels of engagement, we were able to estimate likely response rates among these groups and

tailor the invitee list for the full survey so as to maximise the likelihood of obtaining a representative sample. Thus, all registered respondents in the lowest quartile of engagement were invited to participate in the full survey, followed by diminishing proportions of randomly-selected participants in higher quartiles of classification activity. The result is that the distribution of activity (number of classifications submitted) for the survey sample broadly matches the distribution observed for the whole population of Zooniverse volunteers, as demonstrated below in Figure 2. Although our survey sample under-represents the number of users supplying only a single classification to their respective projects, it does appear to quite closely match the long-tailed distribution of other users supplying two or more classifications. A significant majority of respondents were from the US (39%), followed by the UK (28%), as well as other countries in Western Europe. Although our survey sample showed a slight over-representation of participants from the UK and a slight underrepresentation of respondents from countries outside of the eight most prevalent, the geographical distribution of survey respondents is also broadly representative of the population of Zooniverse users.

[Figure 2 about here]

5. Analysis

5.1. What is the typical profile of an online volunteer?

Descriptive statistics for the respondents to the survey sample are presented below in Table 1. As can be seen from our various measures of engagement, the distribution of voluntary contributions made by Zooniverse users is highly skewed. This can be demonstrated through

a simple analysis of total classification activity, which is a measure of the aggregate data input contributed by each volunteer towards Zooniverse projects (e.g. submission of information relating to a single image counts as a single classification). A raw count of the total number of Zooniverse classifications recorded among the sample shows a mean of around 2,733 per user, versus a standard deviation around 7 times larger than the mean and a median of just 260. A further investigation of the distribution of classification activity among users show that the top 10% of volunteers by overall classification count provide around 80% of the total recorded classifications, the top 5% provide 70%. In other words, the top 5% of contributors supply more than twice as much voluntary effort than the other 95% combined. The extreme level of inequality is highlighted by a Gini coefficient of 0.917 for the cumulative count of classification activity recorded across the entire sample.

[Table 1 about here]

The same skewed distribution of activity is also observed for the number of classifications contributed to a user's 'home' or most frequently visited project (a mean of about 1,811 compared with a median of 162), as well as the total amount of time spent volunteering for Zooniverse projects (a mean of around 29 hours compared with a median of about 3 hours). We observe the same highly-skewed pattern of engagement for our measures of volunteer retention, including the total number of unique sessions and days spent classifying by each contributor. Overall, we can see that a vast majority of volunteers supply a relatively small number of classifications over a very short period of time; usually a handful of sessions lasting only a few hours in total. A majority of the volunteer labour input for these projects is

made by a relatively small minority of volunteers who contribute a greater amount of information over a much longer time period.

A further analysis of the descriptive statistics for the sample shows that the population is reasonably equally divided between males and females (56% male, 44% female) and is mainly composed of white respondents living in cities; just 13% of respondents are non-white and about a third of respondents live in rural communities. The sample also appears to be reasonably affluent; just over half own their own homes, with an average annual income of just over $40,000 per annum. The average educational attainment among the sample is also very high, with typical International Standard Classification of Education (ISCED) qualification levels of around 6 (Bachelor's degree or equivalent). In our dataset, Zooniverse participants reported that around 67% hold a Bachelor's degree or higher. Of these, around 36% hold Master's Degrees and around 12% have doctorate-level qualifications, while around half of the total number of respondents hold these qualifications in science-related subjects.

A visual breakdown of these statistics can be found below in Figure 2, containing histograms of the age, income and education profiles for our full sample of respondents, as well as a comparison of US respondents against the US population based on data obtained from the US Census Bureau (2013). The most significant contrast appears to be in terms of average levels of educational attainment; more than twice the proportion of online volunteers based in the US are educated to degree level or higher compared with the US population, with correspondingly lower proportions educated to a high-school level or lower. This leads us to conclude that our sample of online volunteers is broadly representative of the US population in terms of age and income, but tends to be significantly better educated on average.

[Figure 2 about here]

5.2. What are the motivations for volunteering online?

In addition to collecting socio-demographic information, the survey also collected information on motivations to volunteer, using a subset of questions from the well-known Volunteer Functions Inventory used by Clary *et al.* (1996) and numerous subsequent studies. Volunteers were asked to respond to a subset of three out of the five questions under each heading of the VFI. We selected these subsets of three questions based on those which Clary *et al.* (1998) demonstrate to correlate most strongly with the underlying factor scores for each volunteer motivation. In most cases, the wording for each question needed to be modified only slightly to make it specific to the particular context of online volunteering via the Zooniverse. For example, one of the questions on the VFI under the heading of 'Enhancement' originally reads 'Volunteering makes me feel better about myself'; for the purposes of our survey, this was simply changed to 'Participating in Zooniverse projects makes me feel better about myself'. The one exception to this is the 'Values' motivation categorised by the VFI, which did not seem to be appropriate or applicable given the particular context of this voluntary activity; thus questions originally asking about being 'concerned over others less fortunate than oneself' (or similar) were replaced with questions asking about the extent to which the respondent believes that scientific research benefits society and whether scientific research receives adequate funding. Survey respondents were asked to indicate the extent to which they agreed or disagreed with each of these statements on a 7-point Likert scale and were each presented with the statements in a random order.

A principal component analysis was undertaken using this set of Likert scale data to establish whether the responses reflected the same latent constructs as intended on the VFI. Table 2 (below) summarises the key variables in this analysis as well as the factor loadings or correlation between each individual attitudinal response and the respective factor score. The Kaiser-Meyer-Olkin measure of sampling adequacy is 0.860, indicating that the data is extremely well-suited for principal component analysis. A Varimax rotation method is used with Kaiser normalisation and reveals five latent variables with Eigenvalues in excess of 1, meaning that they each explain more variance in the dataset than any one observed variable taken individually. This is one fewer than the expected six distinct motivations identified by the VFI and occurs due to responses under the 'Protective' and 'Enhancement' categories being identified as being strongly correlated with the same latent construct. Otherwise, each of the other identified factors corresponds clearly and distinctly to the expected items on the VFI scale.

[Table 2 about here]

Given that we are able to group responses into these five underlying factors, we simply aggregate the raw response data on the 7 point Likert scale for each of our identified factors so that the broad trends of responses can be intuitively interpreted. Given that our items each consist of responses to three questions, the aggregated raw scores range from a minimum of 3 to a maximum of 21. Figure 3 contains a visual summary of these raw scores for each item, as well as a summary of the average response within each item (maximum 21), with the averages across the three individual components of each item in parentheses (maximum 7).

Overall, it can be observed that there is a strong skew towards more positive responses for the Understanding item, while equivalent negative skews exist for the Career item, as well as to some extent the Social item. The Protective and Enhancement motivations appear somewhat closer to a normal distribution. To some extent, this is also the case for the Values motivation, although this particular distribution is also fairly leptokurtic, with values quite tightly distributed around the mean value of around 15/21. This indicates that the volunteers who responded to our survey typically appear to be motivated by Understanding and to an extent Values, while Career and Social motivations to not appear to be significant motivators for volunteering in this particular context.

[Figure 3 about here]

5.3. How do different motivations relate to variations in activity and retention among online volunteers?

One limitation associated with analysing the raw Likert-scale values for each of the identified motivational factors is that we do not know how they relate to levels of activity and retention. It is therefore important to establish whether any of these motivations associate with significant variations in contribution levels among our sample of online volunteers. In addition to grouping and reporting the raw response data for each item, we also generate a set of factor scores in each instance using the Bartlett procedure, chosen due to their unbiased estimates of the factor score parameters and high correlation with the estimated factors (DiStefano *et al*., 2009). We use these factor scores as explanatory variables in a regression analysis using observed measures of volunteer activity in model specifications (i) – (iv) and

volunteer retention in model specifications (v) – (vii), as dependent variables. We also use other variables captured by our survey as controls for variations in individual socio-demographic and lifestyle choices across the sample. The results of these regressions are presented in Table 3 (below), with definitions of each variable appearing earlier in Table 1. Although some coefficient estimates are not reported to conserve space, we do include a range of controls for the different online volunteering projects in our sample to control for heterogeneity, particularly with respect to the engagement measures that directly or indirectly relate to classification input as opposed to frequency of visits. This is important to note given that each project involves asking volunteers to undertake a slightly different set of tasks as part of contributing a single data input or classification.

[Table 3 about here]

Although we use a variety of measures of volunteer activity and retention as dependent variables, the broad conclusions are similar across model specifications, indicating that our results are robust to a wide variety of measures of volunteer activity and retention. The most significant positive association between engagement levels and motivation appears to relate to the Understanding motivation, where our coefficient estimates are universally found to be larger than any others for all measures of activity and retention. These regression results therefore show clear evidence that the most active participants in these projects are primarily motivated by a desire to enhance their levels of knowledge and understanding as a result of their participation. We therefore suggest the learning experience of Zooniverse volunteers to be of paramount importance to contributors; projects should offer clear opportunities for

learning to incentivise participation among those groups who are more likely to engage significantly with the platform, recognising that a balance may need to be achieved between learning in ways that are distinct from the main task of classifying. Also, it seems reasonable to suggest that projects should encourage and promote opportunities for learning among all participants given the significant and positive association we find between the Understanding motivation and levels of voluntary input.

We also show evidence of a generally positive association between volunteer participation levels and both the Protective & Enhancement and Values motivations. Although somewhat weaker than the Understanding motivation, we do show evidence that the Protective & Enhancement motivation associates positively and significantly with most measures of activity and retention. This suggests that more committed volunteers do tend to be motivated at least partly by a desire to escape their troubles and/or feel better about themselves as a result of having contributed time and effort towards a worthwhile cause. Interestingly, our measure of (Science) Values, which is cited as being among the leading motivations by Raddick *et al.* (2010; 2013), seems to offer a much stronger explanation for volunteer retention than for activity. In other words, we show that respondents who score higher for the Values motivation are much more likely to actively contribute towards projects over longer time periods.

Conversely, we can clearly see strong negative relationships between both the Career and Social motivations and all of our measures of volunteer activity and retention, suggesting that the most active and committed participants are not motivated by the possibility to enhance their careers or to socialise with other volunteers. While the negative association with the Career motivation might be expected in this context, the lack of social motivation among volunteers for network-dependent online projects may be considered something of a surprise.

Although this may partly be a result of the way in which these projects have been designed (volunteer classifications need to be independent of one another to ensure statistical validity of the findings), the Zooniverse does offer reasonably extensive facilities for interaction and discussion within its community of volunteers. The negative association we find between the Social motivation and each of our measures of engagement may therefore imply some degree of substitutability between social interaction and cognitive input into these online projects. In other words, more committed online volunteers prefer to contribute more intensively to projects than discussing their activities with other participants.

Surprisingly, we show only very limited association between our other socio-demographic controls and either the activity or retention of online volunteers; particularly age, gender, ethnicity and education. The latter finding in particular is significant and suggests that even though the sample of online volunteers seems to be relatively highly educated compared with the rest of the population, we find no evidence of filtering among contributors such that those with the highest education levels contribute the most information. Additionally, income levels do not seem to significantly affect volunteer activity or retention. The only effective constraint on contribution seems to relate to the marital/relationship status of the participant, with significant reductions in the number of classifications supplied in aggregate and for the 'home' project typically observed among those respondents in a committed relationship compared with those who are single. The time control variable (Duration) also indicates that volunteers who have held accounts for longer periods tend to have been more active over time in terms of classification activity and number of visits (retention); a result that is both intuitive and expected.

Altogether, taken across all of our model specifications, the five factor scores generated from the VFI motivation items seem to do a much better job of explaining variations in volunteer

activity and retention than our socio-demographic controls. This indicates that individual level motivations are more powerful predictors of variations in the activity and retention of online volunteers than those reflecting respondent characteristics and lifestyle choices. These findings also have implications for the broader understanding of motivation of this new form of volunteering. In other contexts, it sends to be the case that 'other oriented' motivations (Values, Understanding and Social) dominate 'self-oriented' motivations (Protective, Enhancement and Career) in explaining variations in levels of voluntary engagement; a trend which seems to hold only partially in this context. While some 'other-oriented' motivations are shown to be important (particularly Understanding and to some extent Values), we also show that the self-oriented motivations of Protective & Enhancement are as much or equally important in explaining variations in volunteer activity and retention. We therefore conclude that, while online volunteering appears to demonstrate a degree of commonality other forms of volunteering, the specific combination of motivations we show to relate to volunteer activity and retention in this context suggest that online volunteering to be a new and somewhat distinct phenomenon worthy of special investigation in its own right.

6. Conclusions

This study has introduced a unique and previously underexplored form of volunteering taking place online, which provides opportunities for citizens to engage in research-related activity and analysis for a wide variety of non-profit and charitable organisations. The profile and motivations of these volunteers is explored via direct access to the database of voluntary activity and retention to the portfolio of online volunteering projects hosted by the

Zooniverse and is supplemented by an online survey undertaken with a representative group of volunteers across a representative selection of projects.

Our analysis shows that online volunteers in this context are likely to be relatively affluent and very well educated. We also show evidence that volunteers to these projects are very likely to be white and around twice as likely to live in a city as opposed to a rural area. An analysis of responses to questions adapted from the VFI shows that motivations to participate in these projects can be broken down into five broad categories in line with expectations; Protective & Enhancement, Understanding, Social, Career and Values. An analysis of the raw response data shows that the dominant motivations among the sample are Understanding and Values, indicating that the bulk of participants are motivated by a desire for learning and a positive disposition towards the process of scientific research. Conversely, Career and Social motivations are much less prevalent among this sample than the others, indicating that respondents are less motivated to participate out of a desire for interaction or to gain an advantage in the labour market.

A regression analysis of data relating to observed patterns of engagement shows only limited relationships between key socio-demographic variables and levels of volunteer activity. Although we observe that sample is relatively highly educated, we do not find any significant relationship between education levels and participation rates, which suggests that these projects are not filtering such that the most educated volunteers provide the greatest amounts of classifications. The use of factor scores reflecting VFI motivations shows Understanding motivation has the strongest positive association with voluntary participation, followed by Protective & Enhancement and Values. Conversely, Career and Social motivations are shown to associate negatively with levels of voluntary participation. The strong positive association between the Understanding motivation and all measures of activity and retention

suggests that the most effective incentive that online volunteering projects of this nature can offer is the opportunity for learning. It therefore appears that online volunteering may be more concerned with knowledge creation and human capital enhancement than more traditional forms of 'real-world' volunteering.

References

Agostinho, D., & Paço, A. (2012). Analysis of the motivations, generativity and demographics of the food bank volunteer. *International Journal of Nonprofit and Voluntary Sector Marketing*, 17(3), 249-261.

Allison, L. D., Okun, M. A., & Dutridge, K. S. (2002). Assessing volunteer motives: A comparison of an open-ended probe and Likert rating scales. *Journal of Community & Applied Social Psychology*, *12*(4), 243-255.

Amichai-Hamburger, Y. (2008). Potential and promise of online volunteering. *Computers in Human Behavior*, *24*(2), 544-562.

Bowman, W. (2009). The economic value of volunteers to nonprofit organizations. *Nonprofit Management and Leadership*, *19*(4), 491-506.

Clary, E. G., Snyder, M. & Stukas, A. A. (1996). Volunteers' motivations: Findings from a national survey. *Nonprofit and Voluntary Sector Quarterly*, 25(4), 485-505.

Clary, E. G., Snyder, M., Ridge, R. D., Copeland, J., Stukas, A. A., Haugen, J., & Miene, P. (1998). Understanding and assessing the motivations of volunteers: a functional approach. *Journal of Personality and Social Psychology*, 74(6), 1516.

Clary, E. G., & Snyder, M. (1999). The motivations to volunteer theoretical and practical considerations. *Current Directions in Psychological Science*, 8(5), 156-159.

Cox, J., Oh, E. Y., Simmons, B., Lintott, C., Masters, K., Graham, G., Greenhill, A. & Holmes, K. (2015). How is success defined and measured in online citizen science? A case study of Zooniverse projects. *Computing in Science and Engineering*, 17(4), 28-41.

Dhebar, B. B., & Stokes, B. (2008). A nonprofit manager's guide to online volunteering. *Nonprofit Management and Leadership, 18*(4), 497.

DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation*, 14(20), 1-11.

Fortson, L., Masters, K., Nichol, R., Edmondson, E. M., Lintott, C., Raddick, J., & Wallin, J. (2012). Galaxy Zoo. *Advances in Machine Learning and Data Mining for Astronomy*, 213-223.

Gage, R.L. & Thapa, B. (2012). Volunteer motivations and constraints among college students analysis of the volunteer function inventory and leisure constraints models. *Nonprofit and Voluntary Sector Quarterly*, 41(3), 405-430.

Galton, F. (1907a). Letters to the Editor: The Ballot-Box. Nature 75, 900–1.

Galton, F. (1907b). Vox Populi. Nature 75, 450–1.

Garner, J.T., & Garner, L.T. (2011). Volunteering an Opinion Organizational Voice and Volunteer Retention in Nonprofit Organizations. *Nonprofit and Voluntary Sector Quarterly*, *40*(5), 813-828.

Katz, D. (1960). The functional approach to the study of attitudes. *Public Opinion Quarterly*, 24(2), 163-204.

Lintott, C. J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., ... & Vandenberg, J. (2008). Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society*, *389*(3), 1179-1189.

Lintott, C. J., Schawinski, K., Keel, W., Van Arkel, H., Bennert, N., Edmondson, E., ... & Vandenberg, J. (2009). Galaxy Zoo: 'Hanny's Voorwerp', a quasar light echo? *Monthly Notices of the Royal Astronomical Society*, *399*(1), 129-140.

Misje, A. H., Bosnes, V., Gåsdal, O., & Heier, H. E. (2005). Motivation, recruitment and retention of voluntary non-remunerated blood donors: a survey-based questionnaire study. *Vox Sanguinis*, 89(4), 236-244.

Nov, O. (2007). What motivates wikipedians? *Communications of the ACM*, 50(11), 60-64.

Peachey, J. W., Lyras, A., Cohen, A., Bruening, J. E., & Cunningham, G. B. (2013). Exploring the motives and retention factors of sport-for-development volunteers. *Nonprofit and Voluntary Sector Quarterly*, 0899764013501579.

Planalp, S., & Trost, M. (2009). Motivations of hospice volunteers. *American Journal of Hospice and Palliative Medicine*.

Raddick, M. J., Bracey, G., Gay, P. L., Lintott, C. J., Murray, P., Schawinski, K., & Vandenberg, J. (2010). Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers. *Astronomy Education Review*, *9*(1), 010103.

Raddick, M. J., Bracey, G., Gay, P. L., Lintott, C. J., Cardamone, C., Murray, P., ... & Vandenberg, J. (2013). Galaxy Zoo: Motivations of citizen scientists .*arXiv preprint arXiv:1303.6886*.

Roy, K. & Ziemek, S. (2000). On the economics of volunteering. *ZEF Discussion Papers on Economic Policy*, 26.

Salamon, L. M., & Anheier, H. (1998). The Nonprofit Sector in the Developing World. *The Johns Hopkins University*, Baltimore

Smillie, I. (1995). *The alms bazaar: Altruism under fire-non-profit organizations and international development*. International Development Research Centre (IDRC).

Smith, M. B. (1958). Opinions, Personality, and Political Behavior. *American Political Science Review*, 52(1), 1-17.

Snyder, M., Clary, E. G., & Stukas, A. A. (2000). The functional approach to volunteerism. *Why we Evaluate: Functions of Attitudes*, 365-393.

Stukas, A. A., Worth, K. A., Clary, E. G., & Snyder, M. (2009). The matching of motivations to affordances in the volunteer environment an index for assessing the impact of multiple matches on volunteer outcomes. *Nonprofit and Voluntary Sector Quarterly*, 38(1), 5-28.

Stukas, A. A., Hoye, R., Nicholson, M., Brown, K. M. & Aisbett, L. (2014). Motivations to Volunteer and Their Associations with Volunteers' Well-Being. *Nonprofit and Voluntary Sector Quarterly*, December, 1-21.

Weisbrod, B. (1977). *The voluntary non-profit economy*. Lexington Books: Lexington MA.

Yang, H. L., & Lai, C. Y. (2010). Motivations of Wikipedia content contributors. *Computers in Human Behavior*, *26*(6), 1377-1383.

Tables and Figures

Table 1: Descriptive Statistics

Variable	Description	Mean	Std Dev	Median	Min	Max
Measures of Volunteer Activity						
All Classifications	Total number of classifications completed by respondent across all Zooniverse projects.	2732.73	19876.97	260	1	580,000
Home Classifications	Highest total number of classifications completed by respondent in a given individual project.	1811.24	11989.43	162	1	330,000
Time	Total amount of time (in hours) spent providing classifications across all Zooniverse projects.	29.40	207.75	3.45	0	6,895
Number of Projects	Number of unique projects for which the respondent has recorded at least one classification.	5.81	5.50	4	1	35
Measures of Volunteer Retention						
Sessions	Number of unique log-in sessions recorded across all Zooniverse projects.	43.51	191.62	9	1	6,125
Days	Number of unique days on which the respondent supplied classifications.	29.29	89.00	8	1	2,031
Active Period	Difference (measured in days) between the date of the first and last classifications recorded by the respondent.	841.62	809.71	608	1	2,937
Home Project Controls						
Galaxy Zoo (Base)	Respondent contributes to and answered questions relating to the Galaxy Zoo project.	0.299	-	-	0	1
Planet Hunters	Respondent contributes to and answered questions relating to the Planet Hunters project.	0.247	-	-	0	1
Penguin Watch	Respondent contributes to and answered questions relating to the Penguin Watch project.	0.207	-	-	0	1
Seafloor Explorer	Respondent contributes to and answered questions relating to the Seafloor Explorer project.	0.161	-	-	0	1
Snapshot Serengeti	Respondent contributes to and answered questions relating to the Snapshot Serengeti project.	0.086	-	-	0	1
Other Controls						
Duration	Period of time (in days) between the date of first classification and the date of the survey.	1225.43	788.25	1048	145	2,942
Gender (Female)	Dummy variable if respondent indicated their gender to be female.	0.442	-	-	0	1
Age	Respondent's self-reported age in years.	43.843	15.941	44	18	85
Ethnicity (Non-White)	Dummy variable if respondent indicated their ethnicity to be non-white.	0.129	-	-	0	1
Community Type (Rural)	Dummy variable if respondent indicates they live in a rural area.	0.339	-	-	0	1
Income	Respondent's self-reported income in 2015 USD	41,205	62,541	28,220	0	1,200,000
Religious	Dummy variable if respondent indicated belonging to a religious faith	0.298	-	0	0	1
Charity Donations	Sum of respondent's annual charitable donations in 2015 USD	862.732	2919.201	116.800	0	50,000
Paid Work	Number of hours of paid work undertaken by the respondent in a typical week	23.905	20.569	30	0	95
Relationship Status (Married/Relationship)	Dummy variable if respondent indicates that they are married or involved in a relationship.	0.49	-	-	0	1
Number of children (aged under 12)	Respondent's number of children aged under 12 years.	0.240	0.626	0	0	6
Number of children (aged under 18)	Respondent's number of children aged under 18 years.	0.126	0.424	0	0	4
Number of children (aged 18+)	Respondent's number of children aged over 18 years.	0.602	1.090	0	0	8
Education Level	Highest educational attainment achieved by the respondent (ISCED Category).	6.587	1.689	7	1	9
Parental Education	Highest educational attainment achieved by either of the respondent's parents (ISCED Category).	5.700	2.098	6	1	9
Science Qualifications	Dummy variable reflecting whether the respondent indicated that the highest qualification achieved was in a scientific field.	0.500	-	-	0	1

Table 2: Principal Component Analysis

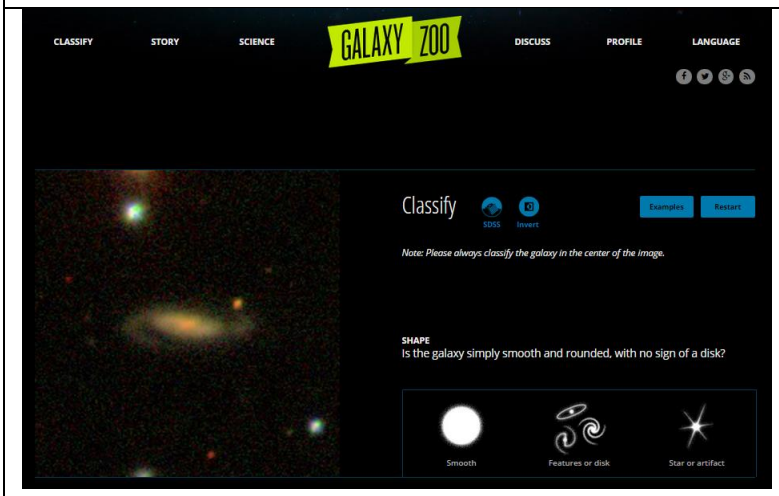
Variable	Factor Loading	Mean	Std Dev	Median	Min	Max
Factor 1: Protective & Enhancement (Eigenvalue = 5.176; 28.75% of Variance Explained)						
Participating in Zooniverse projects offers a good way to escape from my troubles	0.680	3.576	1.717	4	1	7
Participating in Zooniverse projects makes me feel less lonely.	0.668	2.929	1.510	3	1	7
Participating in Zooniverse projects makes me feel less guilty about doing enough to support worthwhile causes.	0.654	3.424	1.651	4	1	7
Participating in Zooniverse projects increases my self-esteem.	0.799	4.068	1.505	4	1	7
Participating in Zooniverse projects makes me feel better about myself.	0.784	4.587	1.438	5	1	7
Participating in Zooniverse projects makes me feel needed.	0.753	4.343	1.573	5	1	7
Factor 2: Understanding (Eigenvalue = 2.062; 11.45% of Variance Explained)						
Participating in Zooniverse projects lets me learn through direct, hands-on experience of scientific research.	0.792	5.508	1.255	6	1	7
I feel the Zooniverse allows me to gain a new perspective on scientific research.	0.870	5.496	1.204	6	1	7
Zooniverse projects help me learn about science.	0.885	5.625	1.116	6	1	7
Factor 3: Social (Eigenvalue = 1.982; 11.01% of Variance Explained)						
Others with whom I am close place a high value on Zooniverse projects.	0.807	2.928	1.496	3	1	7
My friends contribute to Zooniverse projects.	0.804	2.774	1.513	2	1	7
People I know share an interest in Zooniverse projects.	0.857	3.203	1.641	3	1	7
Factor 4: Career (Eigenvalue = 1.411; 7.84% of Variance Explained)						
Participating in Zooniverse projects helps me make new contacts that might help my business or career.	0.813	2.355	1.325	2	1	7
Participating in Zooniverse projects allows me to explore different career options.	0.798	2.881	1.573	2	1	7
Participating in Zooniverse projects will help me to succeed in my chosen profession.	0.744	2.531	1.477	2	1	7
Factor 5: Values (Eigenvalue = 1.054; 5.86% of Variance Explained)						
Participating in Zooniverse projects allows me to support a cause I consider to be important.	0.488	6.045	0.994	6	1	7
Scientific research is adequately funded through government taxation.	-0.752	3.825	2.129	5	1	6
All of society benefits from scientific research.	0.727	6.378	0.979	7	1	7
Kaiser-Meyer-Olkin measure of sampling adequacy = 0.860 Bartlett's test of sphericity = 12,636***						

Table 3: Multiple Regression Analysis

	VOLUNTEER ACTIVITY				VOLUNTEER RETENTION		
	(i) Ln(All Classifications)	(ii) Ln(Home Classifications)	(iii) Ln(Time)	(iv) Ln(Number of Projects)	(v) Ln(Sessions)	(vi) Ln(Days)	(vii) Ln(Active Period)
Protective & Enhancement (Factor Score)	0.099 **	0.073 *	0.102 **	0.065 ***	0.073 **	0.075 ***	0.059
	(0.044)	(0.043)	(0.040)	(0.018)	(0.030)	(0.028)	(0.044)
Social (Factor Score)	-0.119 ***	-0.090 **	-0.112 ***	-0.049 ***	-0.079 ***	-0.070 ***	-0.096 **
	(0.043)	(0.042)	(0.040)	(0.018)	(0.029)	(0.027)	(0.045)
Understanding (Factor Score)	0.162 ***	0.137 ***	0.171 ***	0.091 ***	0.125 ***	0.119 ***	0.106 **
	(0.043)	(0.042)	(0.039)	(0.018)	(0.028)	(0.026)	(0.045)
Career (Factor Score)	-0.208 ***	-0.210 ***	-0.162 ***	-0.031 *	-0.133 ***	-0.124 ***	-0.157 ***
	(0.041)	(0.040)	(0.038)	(0.018)	(0.028)	(0.026)	(0.042)
Values (Factor Score)	0.060 *	0.043	0.047	0.039 ***	0.053 **	0.046 **	0.074 **
	(0.033)	(0.033)	(0.030)	(0.013)	(0.022)	(0.020)	(0.034)
Gender (Female)	0.051	0.036	0.077	0.036	0.001	-0.007	-0.138
	(0.100)	(0.097)	(0.091)	(0.041)	(0.068)	(0.063)	(0.101)
Ln (Age)	-0.015	0.048	0.037	-0.081	0.049	0.010	-0.110
	(0.171)	(0.167)	(0.155)	(0.069)	(0.115)	(0.107)	(0.171)
Education (Self)	-0.020	-0.026	-0.015	-0.001	-0.023	-0.018	-0.018
	(0.033)	(0.033)	(0.030)	(0.014)	(0.023)	(0.021)	(0.034)
Ethnicity (Non-White)	-0.169	-0.150	-0.208	-0.061	-0.155	-0.135	-0.337 **
	(0.145)	(0.141)	(0.130)	(0.059)	(0.097)	(0.091)	(0.156)
Married	-0.246 **	-0.260 **	-0.151	-0.072	-0.114	-0.116	-0.156
	(0.115)	(0.112)	(0.104)	(0.047)	(0.078)	(0.072)	(0.117)
Ln (Income)	0.008	0.009	-0.001	0.003	0.000	0.000	0.010
	(0.013)	(0.013)	(0.013)	(0.005)	(0.009)	(0.008)	(0.013)
Ln (Duration)	1.063 ***	0.988 ***	0.908 ***	0.383 ***	0.889 ***	0.886 ***	2.160 ***
	(0.073)	(0.071)	(0.068)	(0.032)	(0.051)	(0.047)	(0.064)
Constant Term	-1.517 **	-1.589 **	-5.079 ***	-1.026 ***	-3.857 ***	-3.848 ***	-8.985 ***
	(0.721)	(0.709)	(0.654)	(0.303)	(0.489)	(0.454)	(0.729)
F Value	19.180 ***	17.150 ***	15.960 ***	18.920 ***	23.540 ***	26.130 ***	60.550 ***
R-Squared	0.189	0.169	0.168	0.192	0.224	0.241	0.356

Significance: * = 90% level, ** = 95% level, *** = 99% level. Other controls included but not reported are as follows: Home Project Controls (Planet Hunters, Penguin Watch, Seafloor Explorer, Snapshot Serengeti, Residence (City), Education (Parents), Science Qualifications, Ln(Paid Work), Ln(Charity Donations), Religions, Number of Children (<12, <18, 18+).

Figure 1: Volunteer Interfaces for Zooniverse Projects



(i) Galaxy Zoo

Launch:

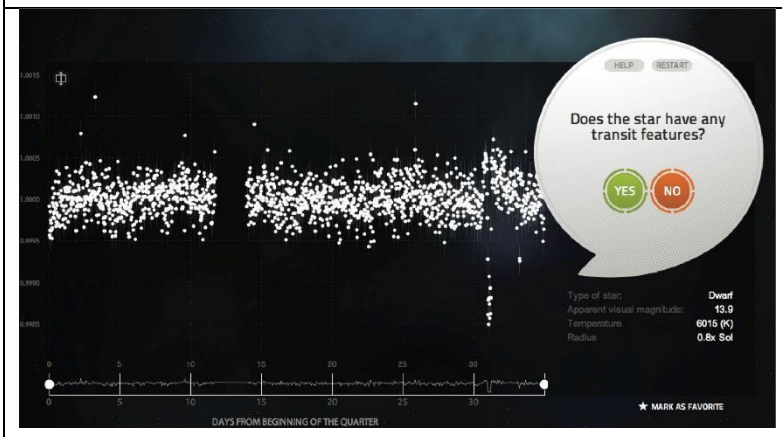
2007 (Galaxy Zoo 1); 2012 (Galaxy Zoo 4)

Number of Registered Volunteers

86,280 (Galaxy Zoo 4)

Brief Description of Task

Answer a series of questions relating to the shapes of deep space galaxies.



(ii) Planet Hunters

Launch

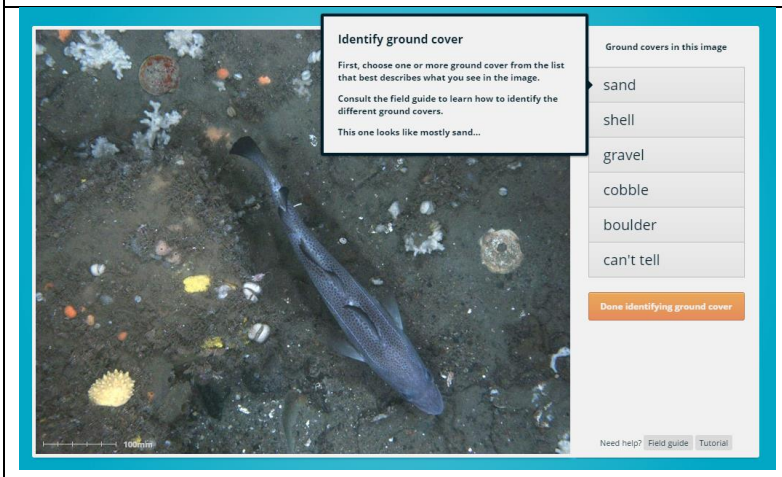
2010

Number of Registered Volunteers

172,628

Brief Description of Task

Identify drops in light that might indicate undiscovered planets passing in front of parent stars.



(iii) Seafloor Explorer

Launch

2012

Number of Registered Volunteers

21,508

Brief Description of Task

Indicate the type of ground cover and the presence, size and shape of marine life in photos of the sea bed.

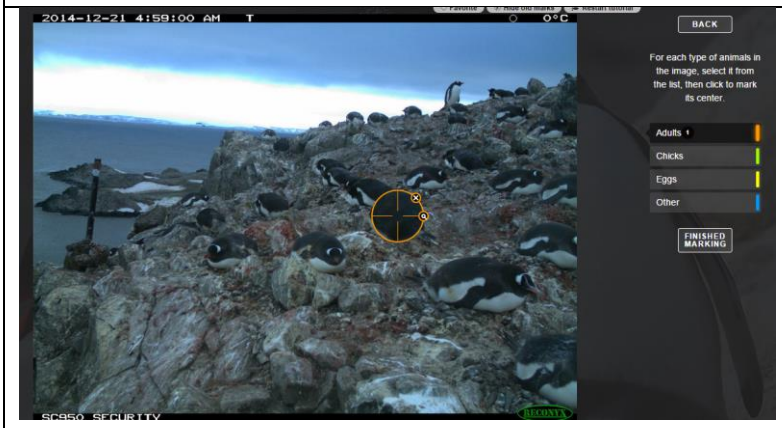
	**(iv) Snapshot Serengeti** Launch 2012 Number of Volunteers 32,429 Brief Description of Task Identify the number and types of animals appearing in images from camera traps on the Serengeti
	**(v) Penguin Watch** Launch 2014 Number of Volunteers 19,499 Brief Description of Task Mark the location and size of penguins appearing in images from the Antarctic.

Figure 2: Distribution of Classification Activity: Zooniverse Population versus Survey Sample

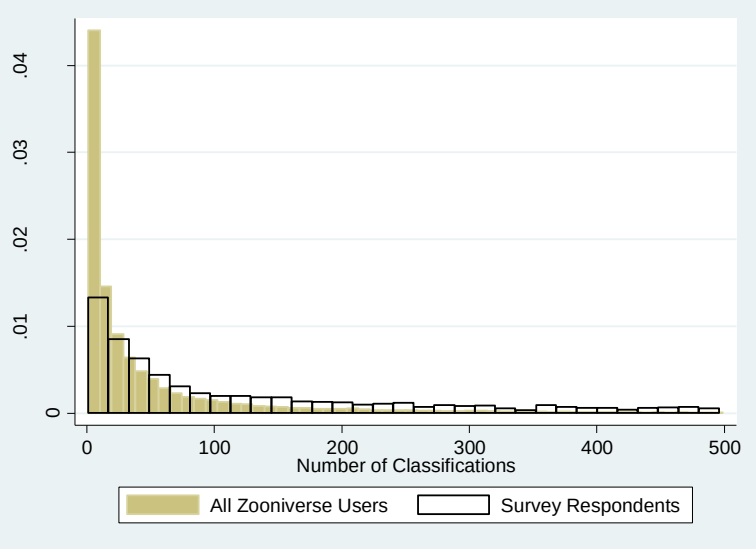
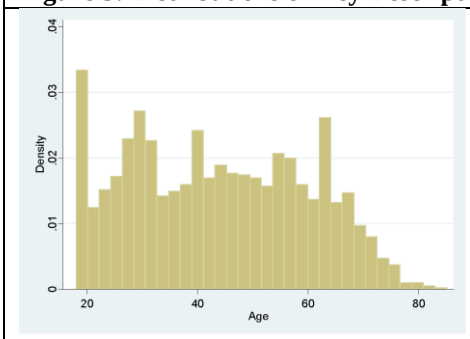
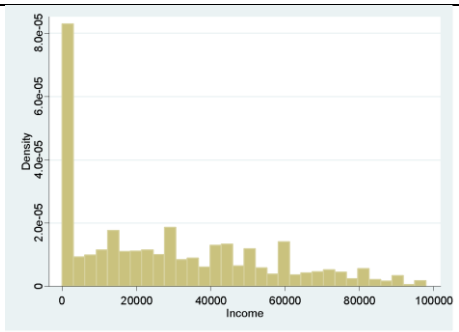


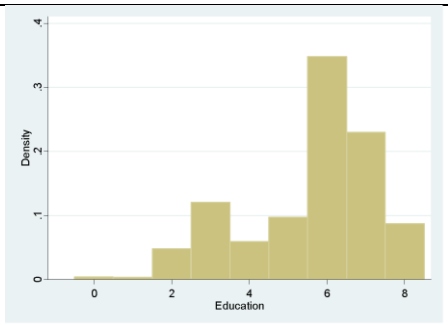
Figure 3: Distributions of Key Descriptive Stats



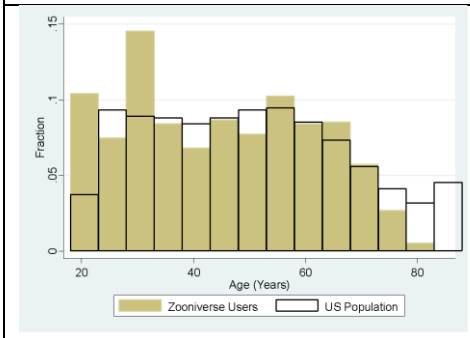
a. Distribution of Respondent Age



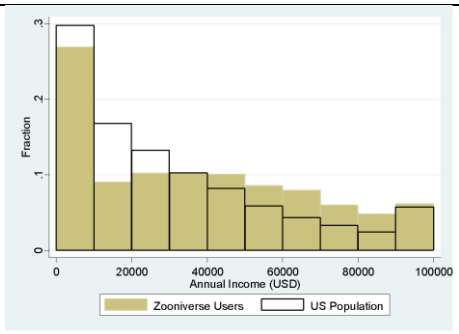
b. Distribution of Respondent Income (USD)



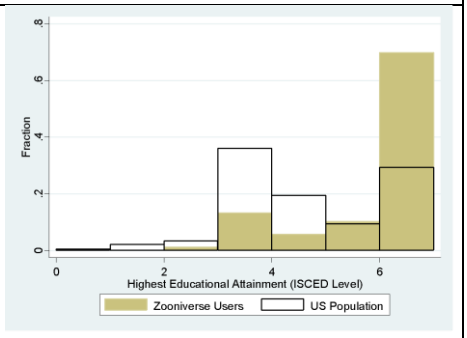
c. Distribution of Respondent Education (ISCED)



i. Comparison of Age Distribution (US Sample versus US Population)

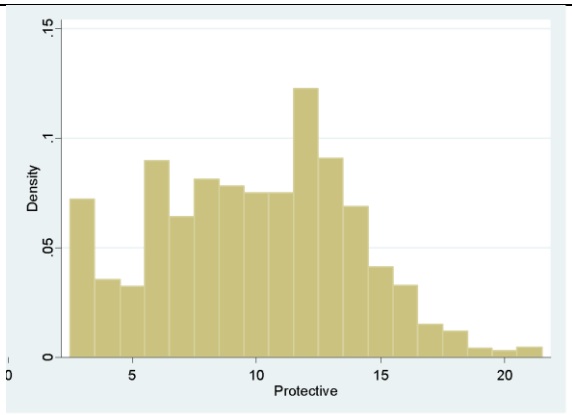


ii. Comparison of Income Distribution (US Sample versus US Population)

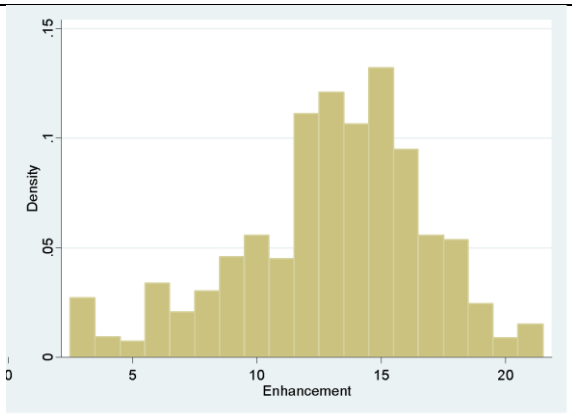


iii. Comparison of Education Distribution (US Sample versus US Population)

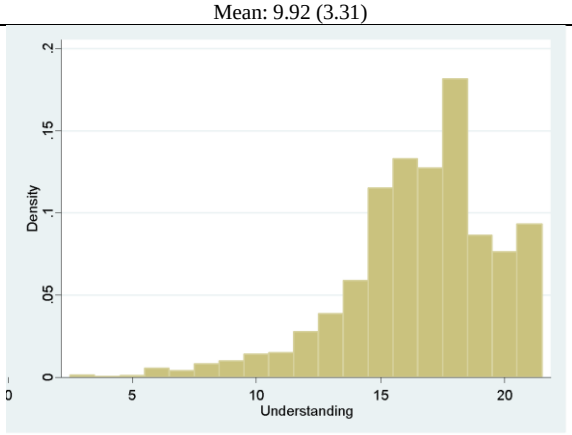
Figure 4: Visual Summary of Raw Responses to VFI Motivation Items



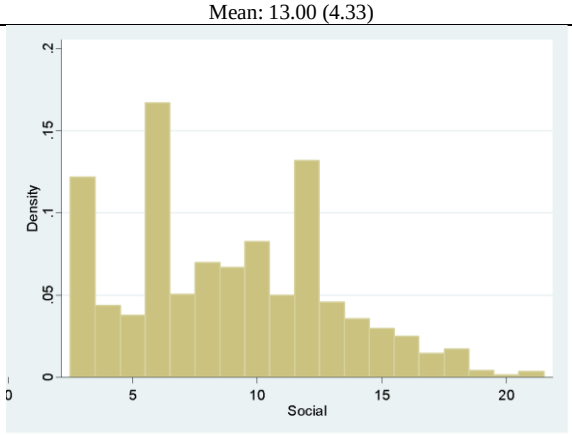
(i.a) Protective
Mean: 9.92 (3.31)



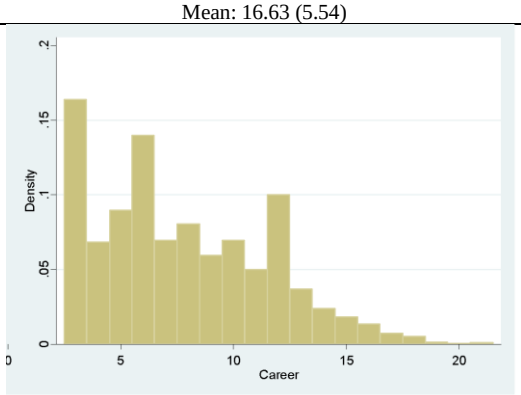
(i.b) Enhancement
Mean: 13.00 (4.33)



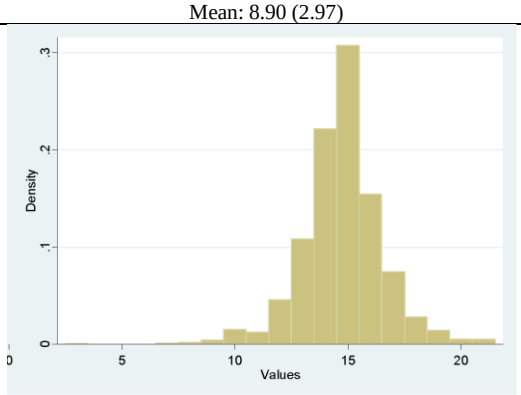
(ii). Understanding
Mean: 16.63 (5.54)



(iii). Social
Mean: 8.90 (2.97)



(iv). Career
Mean: 7.76 (2.59)



(v). Values
Mean: 14.75 (4.92)